面向远程自然交互的人机物三元融合端计算

陈益强 纪雯 钱跃良 朱珍民

摘要：自然的远程交互一直是人们追求的目标。随着普适计算技术和宽频网络技术的发展，在互联网上通过远程视频实现跨地区交流和合作成为可能。但是，传统的视频对话系统视频质量差、操作复杂，影响了远程交互的自然感受。近年来出现的一些远程呈现技术由于专注于提高人与人沟通的质量，而忽视了计算系统正由人机共生系统转化为人机物协调社会所产生的需求，因此难以实现人机物之间的自然远程交互。为了营造自然的沉浸式交互环境，让参与者有身临其境的感觉，本文对人机物三元融合端计算技术进行了研究。三元融合端计算技术是一种以用户为中心实现端内以及端间互动计算的技术，具体包括端内的人-机交互和机物协同以及端间的人-机-人交互、人-机-物交互和物-机-物交互计算技术。其中，端指一个由人、机、物构成的三元微世界。基于人机物三元融合端计算框架，我们搭建了爱心小屋远程亲情互动平台，旨在融合沉浸式人-人交互、启发式人-机交互、高保真机物协同三元交互技术，提供一个面向跨地区交流和合作的沉浸式、易操作、高保真的远程自然交互端平台。目前该平台已经成功在一个打工企业和一个村委会进行一期部署，取得了良好的社会效益。

关键字:远程自然交互　人机物三元融合端计算　沉浸式人-人交互　启发式人-机交互　高保真机物协同

1　需求背景

随着社会快速发展，跨地区交流和合作成为时代的潮流。传统的远程交互需求主要集中在远程会议、远程教育和远程医疗方面。然而，近年来，由于外出务工人员的增多，更人性化的远程亲情互动需求也提上日程。据国家统计局普查，我国外出农民工达 15863 万人，而 56.4%的农民工子女却留守家乡，70%以上的留守儿童一年只能与父母见一次面。科学研究表明，父母与孩子之间缺乏有效沟通会对孩子的身心成长产生不良影响，也不利于父母的心理健康，甚至影响到整个社会的和谐发展。为了改善其成员受到地域阻隔的家庭的交流现状，提升跨地区交流和合作的效率，自然有效的远程交互方式成为社会的重要需求。目前，远程交互主要通过电话和远程视频实现。

由于电话只能呈现远程用户的声音，因此很难实现自然有效的远程交互。现有的视频对话系统则受到设备障碍、眼神交流缺失、环境光照无法控制、网络带宽受限的影响，难以提供跨地区人们之间自然的“面对面”交流体验；另外，由于操作复杂，在用户与设备之间造成了技术壁垒；并且，由于图像分辨率和设备异构的限制，很难实现远程用户之间的物品信息分享和情境融入。因此亟需一种有效地切实改善远程用户之间沟通交流的互动平台。近年来发展的远程视频交互技术，旨在通过真人大小的高清视频和高保真的立体音频，以及全景场景拼接，增加人与人之间远程沟通的沉浸性。但是由于操作复杂，用户很难全身心地关注谈话内容本身；另一方面，由于只考虑到用户的音视频交互，仍然无法满足用户分享物品以及融入和改变对方生活情境的愿望；再次，由于设备昂贵，且需搭建专用网络，无法实现大范围的有效覆盖，难以推广。总体而言，现有的远程交互方式都只是构建在人机共生系统上，只能实现简单的人-机-人交互，而忽视了计算系统正由人机共生系统转化为人机物协调社会的趋势，因此很难满足自然的远程交互需求。

2 远程自然交互系统愿景

理想的远程自然交互系统应该是一种针对人机物协调社会设计的，以人为中心，能够最大程度拉近远程交互双方距离的交互系统。在技术层面上，远程自然交互系统，应该充分考虑人机物三元结构的互动计算技术，以期实现远程交互像在同一个物理空间的面对面交流一样自然的愿景。具体而言，远程自然交互系统应该符合以下特点：（1）系统应满足低成本、易推广的要求，能实现广泛覆盖，有效建立需要进行跨地区办公、教学、医疗以及亲情互动的人们之间的联系；（2）系统要能够提供丰富、逼真的体验效果，能在通用的网络环境状态下实现流畅的高质量的视音频交互，且能让远程用户融入到同一个虚拟环境，并支持对人们的注意力具有重要影响的眼神交流，使得远程交互接近真实场景，让交互双方在互动过程中获得良好的沟通体验，提高沟通的效果；（3）系统要能实现远程物体之间的有效互动，将交互双方的物理世界紧密联系起来，并进行协调控制，保证物理世界的同步和统一；（4）系统要能实现人与远程物体的有效互动，使得人可以融入到远程的物理情境，并通过操控远程物体，改善远程交流的能动性；（5）系统应易于操作，具备良好的人-机交互方式，使得双方的注意力集中在谈话内容本身，而不致受到复杂操作的束缚，系统的交互方式应像人与人之间的交流一样自然；（6）系统要能在互动过程中为双方提供足够的信息，使双方的沟通能够持久、深入，应该满足远程用户之间分享物品，文件材料或生活照片以及娱乐视频的需要，丰富谈话内容，增强趣味性。总之，理想的远程自然交互系统需要充分挖掘人机物三元互动的机理，这对计算模式提出了新的挑战。

3 人机物三元融合端计算

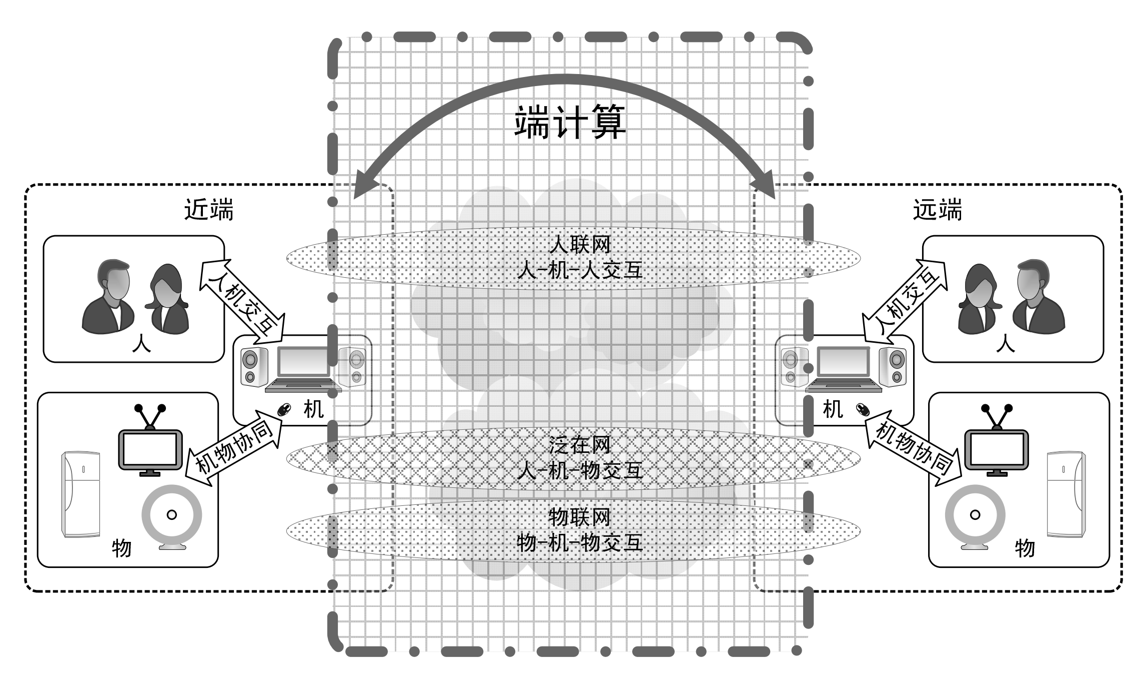


图1. 三元融合端计算示意图

针对远程自然交互系统的特点及其对计算模式提出的挑战，本文提出一种人机物三元融合端计算技术，以期通过计算网络实现人与人之间、物与物之间以及人与物之间的自然远程沟通和有效互动。人机物三元融合端计算技术是一种以用户为中心实现端内和端间互动的计算技术，包括端内的人-机交互和机物协同以及端间的人-机-人交互、物-机-物交互和人-机-

物交互计算技术。其中“端”指一个地方的人、机、物构成的三元世界，如图 1 所示。三元世界是对计算的一种新理解和新的思维模式。区别于以往的一人一机组成的、分工明确的人机共生计算系统，三元世界是由计算世界、物理世界、人类社会组成的人机物协同社会，是一个多人多机多物组成的动态开放协同工作的网络社会。计算系统的变革，要求计算模式也发生新的范式变革，三元计算的概念就此应运而生。三元计算是一种综合利用物理世界、赛博空间（Cyberspace）、人类社会的资源，通过人机物融合合作完成计算任务的计算范式，目的是实现互联网、物联网和社会网的新三网融合，实现信息资源与物理资源、社会资源的有效互动和综合利用。三元融合端计算作为三元计算的一个具体实现，主要用于提供远程的人机物互动。三元计算融入了“人联网”技术，物联网技术和泛在网技术，能够基于个人和社会的需求，实现人与人、人与物、物与物之间在任何地点按需进行的信息获取、传递、存储、认知、决策、使用等功能，具有很强的环境感知、内容感知能力和沉浸性、智能性，为个人和社会提供无所不在的，符合日常生活习惯的信息服务和应用。

4 三元融合端计算示范应用——爱心小屋

4.1 系统概述

针对远程自然亲情互动的需求，基于三元融合端计算技术的理论研究成果，我们搭建了爱心小屋远程亲情互动示范平台，以期加强地域阻隔家庭之间的亲情互动，尤其是农民工与家人之间的亲情互动，疏导、缓解农民工在长期异地工作过程中所形成的情感及心理健康问题，改善留守子女的教育及留守老人的健康，为创新社会管理模式提供一种新的途径。爱心小屋远程亲情互动示范平台是一个基于智能电视机及宽带网络的集成人机物三元结构的远程交流系统。它可以实现简便易操作的、自然状态的远距离可视交流。该系统以三元融合端计算技术为框架，开发了沉浸式人-人交互、启发式人-机交互、高保真机物协同等具体核心技术。其中，人-人交互部分通过对虚实融合技术、自然的眼神交互技术、音频处理技术、以及面向用户体验的流畅传输技术的研究实现跨地区人们之间的沉浸式“面对面”交流；人-机交互部分通过简单自然、符合用户操作习惯的手势操控界面的研究实现人与设备之间的快速交互；机物交互协同部分通过跨设备协同、高质图文共享技术、环境智能协同技术实现跨地区人们之间的信息共享和情感互动。本系统的创新点在于：（1）**应用模式：**本系统是第一个针对农民工及其留守子女和老人等弱势群体的大型社会管理应用；（2）**集成：**本系统是第一个集成人机物三元结构的新型系统；（3）**技术：**本系统融合了沉浸式人-人交互、启发式人-机交互、高保真机物协同技术，其中，远程沉浸式人-人交互技术主要解决“面对面”交流问题，启发式自然人-机交互界面技术主要解决易操作问题，高保真机物协同技术主要解决不同设备的互操作及协同问题。本系统技术发展的总体思路和目标是：致力于三项核心体系及七项核心子技术的研发，并在示范平台上推广应用，以消除地域阻隔，为异地人群营造一个“面对面”的沟通及社交环境，让所有成员自然地身临其境交流，满足现代人对快节奏、个性化、舒适生活的追求。

4.2 主要功能及关键技术

系统的主要功能如图 2 所示。

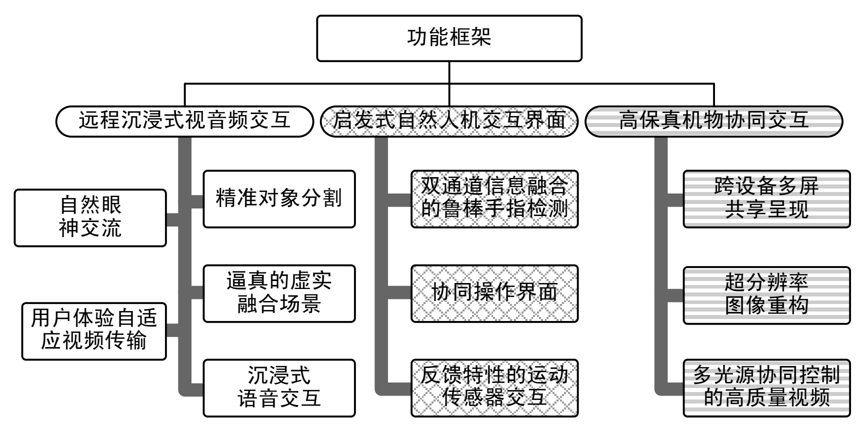


图2.　面向远程沉浸式交互的三元交互终端系统功能图

4.2.1 沉浸式人-人自然交互技术

沉浸式人-人交互技术主要是通过视音频增强处理技术实现自然的远程交流，包括四项子技术，即：（1）基于精准对象分割的虚实融合视频合成技术；（2）基于深度伪三维（3D）信息合成的自然眼神交互技术;（3)面向异构网络和终端的沉浸式视频用户体验自适应技术;（4）沉浸式高保真音频交互技术。

其中基于精准对象分割的虚实融合视频合成技术主要研究在线视频分割技术以及虚实融合技术。在线视频分割技术旨在实时、准确地从在线视频中提取出前景。本技术给出一种基于混合摄像头的分割方法：首先基于深度传感器获取的深度信息实现前景的粗分割，再结合彩色图像提供的颜色、边缘信息给出一种多模决策融合的边界修正算法。虚实融合技术的实质是将计算机制作的虚拟场景与实时分割的前景对象进行数字化的实时合成，使人物与虚拟背景能够天衣无缝地融合，以获得完美的合成画面。本技术还给出一种自适应的摄像头对齐方法，并根据前景距离真实摄像头的深度信息，估算出前景在当前虚拟场景中的缩放比例，使得融合的效果不产生畸变；同时本技术研究一种基于 Lab 空间[1]的光照一致图像合成方法，实现前景图像和虚拟场景图像的真实自然融合，使得远程交互的双方能够融入到统一的虚拟环境，创造一种“在一起”的交互体验。

基于深度伪三维信息合成的自然眼神交互技术，旨在对偏离摄像头的视线进行矫正，以达到直视的效果。在传统的远程视频交互系统中，摄像头一般置于视频屏幕上方，在本地和远程视频之间的双向眼神交互难以实现。为了实现自然眼神交流，本技术提出一种普适化的基于虚拟视角的视线矫正方法。该视线矫正方法设计了一个能自适应用户不同位置的虚拟坐标系及基于此虚拟坐标系的几何模型，通过将实际坐标系下的三维点云数据转换到虚拟坐标系下，并重投影到二维虚拟成像平面，达到矫正头部和视线的效果，使对方感到说话者的视线是向着自己的。

面向异构网络和终端的沉浸式视频用户体验自适应技术主要包括面向异构终端的沉浸式视频用户体验质量模型和非对称带宽的多通道自适应视频传输技术。用户体验质量（QoE，Quality of Experience）定义为用户可以感知的服务质量。本技术针对以异构为特点的复杂的网络环境给出一种用户体验质量的评价模型。该模型从用户角度，将异构网络中多域信息分解为四个关键属性维度：可用性、会话质量、服务延迟和安全性，统称为关键质量指标。多通道视频传输就是在传输过程中利用多条传输通道来传输一路或者几路视频信息的方法。为

[1] Lab color space，一种互补色空间，具有三个维度：亮度（***L***）及 ***a***，***b*** 两个互补色维度

了补偿多通道视频传输资源的小时间尺度的波动，本技术采用实时的视频/音频/多业务自适应平滑系统，以便在最多的视频/音频/多业务数据自适应中获得最好的传输质量。

沉浸式高保真音频交互技术主要研究基于远距离麦克风或麦克风阵列的语音采集及处理和基于麦克风阵列的说话人定位以及音频场景建模和还原技术。本技术采用远距离麦克风或麦克风阵列采集语音，避免用户手持或佩戴麦克风以及传递话筒或开关话筒的额外操作，实现自由、免操作的用户体验。同时，研究语音增强、回声消除、自动增益等语音处理算法，使待传输的语音清晰，音量适中。另外，本技术采用麦克风阵列确定说话人的方向和距离，并结合声音信号传播模型，为真实音频场景建模，计算出一组模型参数，将其随一路音频信号共同传输至远程端；在远程端再将模型参数还原，采用多个音频输出设备产生具有位置感的沉浸式音频。

4.2.2 启发式人-机交互技术

启发式人-机交互技术主要是通过启发式操作界面以及基于异构传感器和反馈特性的手势交互技术实现自然的人-机交互。主要包含三项子技术：（1）启发式操作界面自适应技术；（2）基于双通道深度信息建模的手指精确检测技术；（3）基于具有反馈特性的运动传感器交互技术。

由于农民工、留守老人和儿童群体往往对计算机等新技术具有陌生感和排斥感，面对鼠标、键盘束手无策。为了能够将现代化信息技术造福于这一特殊的群体、使他们彼此之间进行的跨地域感情交流更加方便易行，本技术通过引入启发式自适应学习算法和利用已经具备的用户背景知识，研究启发式键盘手势输入技术、用户意图隐状态感知学习技术和操作界面自适应调整技术，来实现能够针对特定人进行自适应调整的启发式操作界面。

在启发式操作界面中，手指检测是技术关键之一。现有的手指检测算法大多是针对二维摄像头采集的 RGB[2]数据，容易受到背景颜色和肤色的影响，并且要求用户的手掌要尽量伸直张开且与二维摄像头的采集方向垂直，这影响了人-机交互的自然性。因此，本技术利用双三维摄像头，研发双三维摄像头坐标系与现实世界坐标系的对齐方法、双三维摄像头深度信息的协同采集与融合方法、基于球体模型的人手分割方法、以及基于双通道深度信息建模的手指精确检测方法，使手指识别能够不受背景颜色、肤色、手掌形状和方向的影响，以增强人-机交互的自然性和用户的体验感。

为了在保持较高识别精度的同时增强平台的交互性和用户真实的体验感，本技术将集成基于加速度计和陀螺仪的运动传感器，并使其具有振动、力反馈、放电等反馈特性；研究多源反馈数据的协同感知与处理技术和多源反馈数据的融合决策机制，使用户同时具有不同的触感，以进一步增强用户真实的体验感。同时，为了实现对平台设备的控制和对信息的选择，本技术将研究基于多源反馈数据的手势识别技术，用于实现对用户手势动作的精确识别。

4.2.3 高保真机物协同技术

高保真机物协同技术主要是通过跨设备互联，高清图文共享和环境协同控制技术实现丰富的远程交流。主要包含三项子技术：（1）跨设备多屏共享呈现技术；（2）基于视频序列的超分辨率图像重构技术；（3）面向视频序列的多光源协同控制技术。

跨设备多屏共享呈现技术主要包括多协议语义互译与设备互联、资源共享呈现系统以及基于混合图像编码的远程屏幕共享技术。多协议语义互译与设备互联技术主要用于实现针对

[2] 红绿蓝三原色信号

网络电视、摄像头和智能手机的基本互联功能和扩展互联功能。基本互联功能实现基于 Wifi 和蓝牙（Bluetooth）协议的语义互译，支持 Wifi 和蓝牙设备互联互操作；扩展互联功能将支持更为广泛的 IGRS/UPnP 协议的语义互译，针对远程互动平台的功能扩展，支持更广泛的设备互联互操作。资源共享呈现系统主要完成沉浸式交互与资源共享的同步响应机制、沉浸式交互与共享资源呈现系统的 GUI[3]融合机制。基于混合图像编码的远程屏幕共享将研究实现高清混合图像编码算法、计算机桌面图像序列编码算法、高效截屏技术和消除编码效应的后处理技术，实现高效截屏和屏幕共享。

基于视频序列的超分辨率图像重构技术是一个图像序列重建的处理过程，具体描述如下：如果我们在不同条件下拍摄得到几幅具有相同场景的模糊且有噪声的低分辨率图像，且这些同一场景的多张图片均可使用，每一帧相对于所选择的参考帧都会有位移，在这种条件下将它们集中进行融合处理，使其合成一幅或多幅高品质的超分辨率图像（即分辨率高于原始图像），所重建的结果与任何一幅原始输入图像相比，噪音和图像模糊的现象都减少了，从而可以获取更多原始场景的细节。本技术中的超分辨率重建方法主要包括以下三个环节：（1）运动估计（对低分辨率图像序列进行运动估计，得出帧与帧之间的运动偏移关系）、（2）插值重建（利用运动估计得到的配准参数重建图像）、（3）去模糊去噪，最终得出所估计的超分辨率图像。

面向视频质量的多光源协同控制技术旨在对采集视频进行实时分析，进而协同控制前景光源，实现环境光的平衡，以提高交互视频的质量，增强用户体验。主要包括（1）基于视频分析的环境光平衡度评估方法：利用远程亲情互动平台实时采集的用户视频，从中提取用户的人脸图像，并根据人脸图像左、右两侧的灰度值来估计环境光照强度，并据此计算环境光的平衡度；（2）基于环境光平衡度的多光源协同控制策略：根据评估得到的环境光平衡度，结合前景光源各亮度等级的光照强度，通过使环境光差异度函数最小化来估计各前景光源的合适等级，并据此对前景光源进行控制。

4.3 典型应用场景

“爱心小屋[4]”远程亲情互动系统集成了人机物三元结构，并融合了多项技术成果，能够很好地满足远程交互需求，可应用到多个场景，诸如全家一起过年、留守儿童远程教育和留守老人远程医疗等。

4.3.1 全家一起过年

“全家一起过年”是爱心小屋示范应用系统一个重要的应用场景。由于工作繁忙或春运紧张，在异地工作的农民工经常不能与家人团聚。此时，可以通过在线视频分割技术和视频合成技术将外出人员与家人融入到同一张饭桌前，让他们在不同地点共同感受共聚年夜饭的团圆，举杯同庆，互赠祝福，分享喜悦，并可以通过合影留念，记住美好瞬间；同时，通过眼神交互技术使得每个家庭成员感受到其他亲人的关注，解决他们的情感障碍问题。而面向用户体验的自适应传输技术犹如稳固的桥墩，为这份虚拟的沟通桥梁提供坚实的后盾。另外，高保真的音频交互技术将使得家庭成员可以如面对面交流一样开怀畅谈，启发式的人-机交互技术将减少家庭成员因技术知识的匮乏引发的手足无措的操作，而高保真的机物协同技术将使异地阻隔的家庭成员方便地分享手机中拍摄的报平安的微视频以及送祝福的精美贺卡。

4.3.2 留守儿童远程教育

[3] Graphics User Interface，图形用户界面

[4] “爱心小屋”是基于本文所述技术的平台示范项目

对留守儿童的远程教育是爱心小屋示范应用系统另一个重要的应用场景。主要是通过利用城市或繁华地区的教育资源，采用远程交流的方式，对留守儿童进行综合教育，培养留守儿童自信、自强和合作的人生态度。通过面向用户体验的自适应传输技术可以实现远程教师和孩子之间无障碍的交流，让留守儿童有机会接受教育辅导，提高他们的综合素质和能力。视线矫正技术可以让孩子感受到教师关切的眼神，增强孩子学习的积极性和主动性，实时准确的虚实融合技术可以让教师和孩子置身各种虚拟的学习和游戏场景，让他们犹如在同一个场景互动，促进教师和孩子之间的交流。另外，孩子们还可以通过高清图文共享技术传递自己的作业，提交给教师查阅和指正。

4.3.3 留守老人远程医疗

对留守老人的远程医疗是爱心小屋示范应用系统的又一个重要应用场景。由于农村医疗条件落后以及老人自身行动不便，当留守老人的身体出现异常时，很难及时就医。此时，通过面向用户体验的自适应传输技术提供的高质量视频，可以快速实现远程的医疗咨询和诊断，并能通过实时视频演示以零学习的方式让老人接受科学保健。辅之虚实融合技术还可以让病房变换成家庭住所，避免老人对医院的恐慌，使他们能够更自然地向医生阐述他们的身体状况。同时，视线矫正技术可以唤起老人表达的欲望，促进老人和医生之间的有效沟通。另外，老人还可以通过手机录制他们的生活状况，并采用跨设备互联技术，将这些信息传输给医生，以详尽地描述病情，为医生的诊断提供事实根据。

4.4 研发成果及部署

经过对三元融合端计算技术的深入研究，我们在人-人交互、人-机交互和机-物协同等方面实现了系列理论创新和技术突破，在人-人交互功能部分，主要实现了一种基于混合摄像头的精准、鲁棒的在线视频分割方法，并将其应用到 “我的快照”和“远程合拍”两个系统功能；同时实现了一种用户意图驱动的视频合成方法，并应用于“远程合拍”系统功能；实现了一种基于虚拟视角的视线矫正方法，可提供自然的眼神交互；提出了一种基于分辨率和帧速率的自适应传输方法，可保证异构网络环境下的流畅视频传输；采用远距离麦克风，实现了语音采集和语音增强、自动增益等音频处理功能。在人-机交互功能部分，实现了一种九宫格式的启发式自适应输入界面，同时给出了一种基于双通道球模型的精确手指检测算法以及基于手指检测的手势追踪与识别算法，并成功应用于“图片浏览控制”系统功能，可轻松实现图片的浏览。在机物协同功能部分，主要给出了一种基于二叉树预测的图像编解码方法，通过对操作台上的采集图像进行编码、传输和解码，实现了跨设备（电视、智能手机）图像远程重定向功能；同时对采集的低分辨率图像序列进行超分辨率重建，实现了高清图文共享功能；另外，给出了一种基于视频质量的灯光控制算法，实现了多光源的协同控制功能。基于上述技术成果，我们搭建了爱心小屋远程亲情互动系统的示范平台，并在郑州打工企业和河南生产力促进中心进行了一厂一村试点。部署在郑州打工企业样板间员工宿舍，经来自打工企业 19 个园区、15 个事业群的技能之星，累计百多人次的打工企业员工及河南政府人员试用，获得较好评价。图 3 显示了两个地方的实地部署情况。



图3. 打工企业和河南生产力促进中心部署图

5 结束语

本文介绍了远程自然交互系统的背景及需求，展望了远程自然交互系统的愿景，介绍了面向远程自然交互的人机物三元融合端计算技术，并详细描述了该技术的应用示范平台—爱心小屋远程亲情互动系统的技术构成和应用情况。人机物三元融合端计算技术的研究仍处于初级阶段，我们将在以后进一步对该技术进行深入研究和探讨，以实现跨地区的自然远程交流，使得更多的美好故事走入远程家庭和远程办公人员的生活，为他们的身心健康及工作效率带来福音。

作者简介：

陈益强： 中国科学院计算技术研究所普适计算研究中心，主任、研究员 yqchen@ict.ac.cn

纪 雯： 中国科学院计算技术研究所普适计算研究中心，副研究员 jiwen@ict.ac.cn

钱跃良： 中国科学院计算技术研究所普适计算研究中心，正研级高级工程师、中国科学院计算技术研究所智能技术研究部总工程师

朱珍民： 中国科学院计算技术研究所普适计算研究中心，正研级高级工程师

（ 上接第35页 ）

[25] N.-Y. Liang, G.-B. Huang, P. Saratchandran and N. Sundararajan. A fast and accurate on-line sequential learning algorithm for feedforward networks [J]. IEEE Transactions on Neural Networks, 2006, 17(6): 1411-1423.

作者简介：

于汉超： 中国科学院计算技术研究所普适计算研究中心，北京市移动计算与新型终端重点实验室，博士研究生，yuhanchao@ict.ac.cn

陈益强： 中国科学院计算技术研究所普适计算研究中心，北京市移动计算与新型终端重点实验室，研究员

刘军发： 中国科学院计算技术研究所普适计算研究中心，北京市移动计算与新型终端重点实验室，副研究员

纪 雯： 中国科学院计算技术研究所普适计算研究中心，北京市移动计算与新型终端重点实验室，副研究员